

基于跨语言图神经网络模型的属性级情感分类*

鲍小异, 姜晓彤, 王中卿, 周国栋

(苏州大学 计算机科学与技术学院, 江苏 苏州 215006)

通信作者: 王中卿, E-mail: wangzq@suda.edu.cn



摘要: 目前, 在属性级情感分类任务上较为成熟的有标注数据集均为英文数据集, 而有标注的中文数据集较少. 为了能够更好地利用规模庞大但却缺乏成熟标注数据的中文语言数据集, 针对跨语言属性级情感分类任务进行了研究. 在跨语言属性级情感分类中, 一个核心问题为如何构建不同语言的文本之间的联系. 针对该问题, 在传统的单语言情感分类模型的基础上, 使用图神经网络模型对跨语言词-词、词-句之间的关系信息进行建模, 从而有效地刻画两种语言数据集之间的联系. 通过构建单语词-句之间的联系和双语词-句之间的联系, 将不同语言的文本关联起来, 并利用图神经网络进行建模, 从而实现利用英文数据集预测中文数据集的跨语言神经网络模型. 实验结果表明: 相较于其他基线模型, 所提出的模型在 $F1$ 值指标上有着较大的提升, 从而说明使用图神经网络建立的模型能够有效地应用于跨语言的应用场.

关键词: 图神经网络; 属性级情感分析; 跨语言
中图法分类号: TP18

中文引用格式: 鲍小异, 姜晓彤, 王中卿, 周国栋. 基于跨语言图神经网络模型的属性级情感分类. 软件学报, 2023, 34(2): 676-689. <http://www.jos.org.cn/1000-9825/6667.htm>

英文引用格式: Bao XY, Jiang XT, Wang ZQ, Zhou GD. Cross-lingual Aspect-level Sentiment Classification with Graph Neural Network. Ruan Jian Xue Bao/Journal of Software, 2023, 34(2): 676-689 (in Chinese). <http://www.jos.org.cn/1000-9825/6667.htm>

Cross-lingual Aspect-level Sentiment Classification with Graph Neural Network

BAO Xiao-Yi, JIANG Xiao-Tong, WANG Zhong-Qing, ZHOU Guo-Dong

(School of Computer Science and Technology, Soochow University, Suzhou 215006, China)

Abstract: Most of the mature labeled dataset of aspect-level sentiment analysis are in English, it is quite rare in some low-resource language such as Chinese. For the sake of utilizing the vast but unlabeled Chinese aspect-level sentiment classification dataset, this study works on cross-lingual aspect-level sentiment classification. Nevertheless, the most central and difficult problem in cross-lingual mission is how to construct the connection between the documents in two languages. In order to solve this problem, this study proposes a method using graph neural network structure to model the connection of multilingual word-to-document and word-to-word, which could effectively model the interaction between the high-resource language (source language) and low-resource language (target language). The connections include multilingual word-to-document connection and monolingual word-to-document connection are constructed to tie the source language data and target language data, which are modeled by graph neural network to realize using English labeled dataset as trainset to predict Chinese dataset. Compared with other baseline model, the proposed model achieves a higher performance in $F1$ -score, which indicates that the presented work does contributing to the cross-lingual aspect-level sentiment classification.

Key words: graph neural network; aspect level sentiment analysis; cross-lingual classification

不同于常见的句子级情感分类, 属性级情感分析任务聚焦于长文本中对某个具体属性的情感倾向判断. 例如, “手机的性能不错, 但是电池续航不太好”中就包含了“性能”和“续航”这两个具有相反情感倾向的属性,

* 基金项目: 国家自然科学基金(62076175, 61976146); 江苏省双创博士计划

收稿时间: 2021-10-08; 修改时间: 2022-01-09; 采用时间: 2022-03-05

如果像传统的任务那样局限于句子级的情感, 就会对句子中包含的信息造成极大的浪费. 目前的主流深度学习模型已经在属性级情感分类上取得了不错的性能表现, 如 TD-LSTM^[1]、TC-LSTM^[1]、ATAE-LSTM^[2]、机器阅读理解模型等. 然而, 目前的成就主要依靠于个别单语言范围内大量成熟的有标注数据, 比如英文的 SemEval14 Restaurant/Laptop Review 数据集、Twitter 数据集等. 在许多像中文一样缺乏成熟标注数据集的语言上, 上述单语言模型很难达到理想的性能. 这种不同语言之间的资源不均衡, 使得许多具有大量丰富信息但是未被标注的数据被浪费掉. 基于上述情况, 跨语言的属性级情感分类任务被提出, 其主要目的是使用标注数据资源丰富的源语言数据集作为训练集来训练模型, 将缺乏成熟标注数据的目标语言数据集作为测试集来对模型性能进行评估, 以期提升模型在标注资源匮乏的目标语言上的分类性能表现(如图 1 所示), 最终达成使用跨语言模型对目标语言数据进行标注利用的目的, 从而更好地利用规模庞大但却缺乏成熟标注数据的中文语言数据集. 其中, 最为困难的核心问题为如何构建不同语言的文本之间的联系. 为此, 我们提出了聚焦于词级别关系的跨语言图模型来解决上述问题.

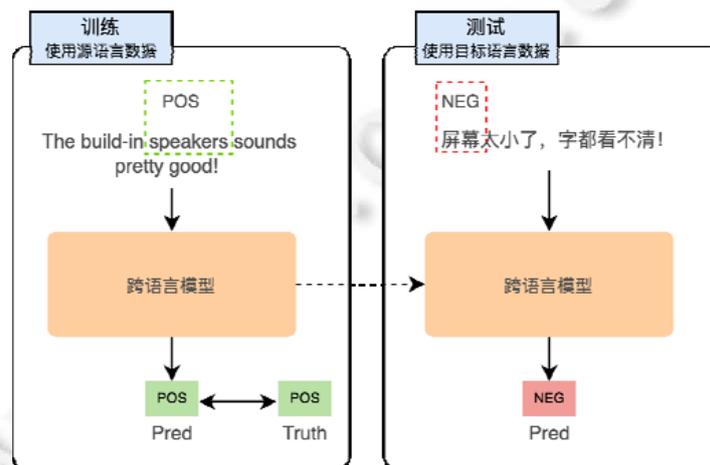


图 1 跨语言属性级情感分类任务示例图

已有一些研究着手解决跨语言的文本分类任务, 比如使用双语词典、双语词向量等等. 在预训练语言模型兴起后, 主流做法开始转向使用多语言预训练深度学习模型(mPLM)^[3]: 使用大量包含多语言数据的语料库来对基于 Transformer 结构的深度学习模型进行预训练, 并在一些下游任务上取得了不错的性能表现. 此外, Wang 等人^[4]提出的 CLHG 模型使用了类似于 TextGCN^[5]的文本图模型, 克服了预训练深度学习模型只注重语义信息而忽视句法信息这一缺点. 但在上述方法中, 无论是 mPLM 还是 CLHG, 均只是聚焦于构建句子级的关系, 忽略了具有丰富信息的词级别的跨语言语义联系, 而后者在属性级情感分析这类关注属性级而非句子级信息的任务上显得更为重要.

为了克服上述问题, 利用词级别信息更好地构建不同语言的文本之间的联系, 我们提出了一种跨语言图模型. 首先, 为了更好地提取双语间的语义联系, 构建边密集图, 我们使用机器翻译将文本语言进行互相转换, 使用源语言以及源语言的翻译去预测目标语言与目标语言的翻译. 我们使用边密集的图模型来构建双语的预训练表征之间的联系, 将源语言和目标语言以及各自翻译的词和句均抽象为图中的节点. 针对属性级情感分类任务的特点, 我们将双语的语义关系细分到词一级, 不仅构建了单语言词-句共现边, 还构建了双语言词-词翻译边、双语言词-句翻译边. 此外, 在最后输出预测结果时, 使用预测目标在目标语言与源语言这两种语言形式的预测概率联合判断结果, 相较于单语言概率判断又有了一定的提升.

综合来说, 我们工作的贡献可以总结为:

- 提出了一种基于图神经网络的跨语言属性级情感分类模型, 并首次在跨语言的单词级别上设计了包

括单语词-句、双语词-句、双语词-词等多种边来构建不同语言文本之间的关联;

- 相较于传统使用单语言数据概率判断分类结果的方法,我们首次提出了同时使用目标语言和目标语言翻译两种语言数据分类概率联合判断分类结果的方法,并与仅使用单语言分类相比,性能上有了提升;
- 进行了一系列可扩展的实验,包括在跨中英文两个数据集的多个领域上的实验以及逆向使用目标语言预测源语言属性级情感倾向的实验.实验结果表明:相较于其他基线模型,我们提出的模型在 $F1$ 值指标上有着较大的提升.

1 相关工作

早期的属性级情感分类主要依靠于复杂的特征工程,通过依照经验人工制定规则的方式,训练一个传统的机器学习分类器.近年来,随着深度学习的发展,有许多工作致力于提升深度学习模型在属性级情感分类上的性能表现:TD-LSTM^[1]和 TC-LSTM^[1]将属性词和上下文区分开,在序列处理的每一个阶段都考虑属性信息.Wang 等人^[2]提出的 ATAE-LSTM 则同样将属性信息与上下文信息拼接在一起,通过编码提取表征,并在序列输出中结合属性特征表示来计算注意力权重.在预训练语言模型,例如 BERT^[3]诞生后,也有许多工作将预训练语言模型应用到属性级情感分析的任务上.Song 等人提出的 AEN-BERT^[6]使用 BERT 分别提取属性词和上下文的编码,并使用注意力机制来构建属性词和上下文之间的语义联系.Zeng 等人提出的 LCF-BERT^[7]则使用 Context Dynamic Weighting/Context Dynamic Mask 的方式,在上下文编码中引入属性词的信息.Mao 等人^[8]使用 BERT 预训练模型作为主干,将机器阅读理解(MRC)的方法应用在属性级情感分析任务上,并取得了 SOTA 的性能表现.因而,我们也选择机器阅读理解模型作为我们的单语言属性级情感分类模型,并以此为基础,加入图模型来构建跨语言联系.

而在跨语言分类方面,主要的思路包括两个方面:建立跨语言文本联系或者使用除去标注方法之外的方法来解决目标语言资源匮乏的问题.对于前者,传统的做法是对语料进行翻译^[9],或者在模型外使用一些双语对齐的工具^[10],例如双语词典^[11-13]或者近似词库等等.近年来的主流做法则是基于跨语言文档训练一些跨语言的词向量,包括双语词向量^[14-16]、Zhang 等人提出的 Common SubWord Representations^[17].大规模预训练语言模型兴起后,出现了许多多语言预训练语言模型(mPLM)^[18],诸如 mBert^[3]、XLM^[19]、XLM-R^[20].其中,mBert 使用共享词汇表的 104 种语言来训练模型,但与单语言 BERT 相比,其模型本身并没有针对多语言进行优化:大多数词汇没有在语言间共享,因此能学到的跨语言知识是很有限的.针对上述问题,XLM 的模型输入改用了字节对编码(BPE),把输入按所有语言中最常见的词片段(sub-words)进行切分,以此来增加跨语言共享的词汇.此外,XLM 还提出了翻译语言模型(TLM):在训练语料中加入含义相同语言不同的训练样本和位置编码,训练中,用一个语言的上下文信息去预测另一个语言被遮住的 token.训练时则同时训练掩码语言模型(MLM)和 TLM.XLM-R 则在结合 XLM 和 RoBERTa^[21]中使用方法的基础上,增加了新模型的语言数量和训练示例的数量,用超过 2 TB 的已经过清理和过滤的数据,以自我监督(self-supervised)的方式训练跨语言表示,实现了在跨语言下游任务上的 SOTA 性能,成为了首个性能上超过单语言 BERT 的多语言预训练模型.此外,与上述使用跨语言的预训练模型或者跨语言词库的思路来预测 low-resource 目标语言不同,Fei 等人提出的 MVEC^[22]则提出了 zero-resource 的概念,强调在完全没有对齐工具或者双语词库的情况下进行跨语言分类任务:通过在模型中加入使用编码器-解码器结构(encoder-decoder)的无监督学习机器翻译(unsupervised machine translation)模型,并在训练时一起使用机器翻译和分类的 Loss 对模型的参数进行调整,从而实现在只有源语言和目标语言各自单语言词库的 zero-resource 情况下进行跨语言文本分类.而面对目标语言标注资源匮乏的问题,Chen 等人^[23]则从文档级别的情感信息入手,利用文档级别信息标注非常容易获得而属性级信息标注非常难以获得的特点,使用胶囊网络(capsule network)将文档级的情感信息加入到模型中以提高在属性级情感分类任务上的性能.Zhang 等人^[17]的思路则与之相反,摒弃了句子级信息转而聚焦于更加细分的字符级信息,提出了 classification aided by convergent orthography (CACO)框架,使用一个字符级别的跨语言编码器对目标语

言和源语言进行编码,再投入到一个词级别的分类器对文档进行分类。

图神经网络自提出后就引起了广泛的关注,并被应用到许多情感分析和文本分类任务当中^[24-28],其中,TextGCN^[5]提出了一种通过将文档和单词都视作图中节点的方法来将文本抽象成图。但是这些工作大多都局限于单语言范围内,只有Li等人^[29]和Wang等人^[4]在跨语言工作上使用了图神经网络,而这两者中:前者只是将图神经网络应用在元学习方面,后者在跨语言方面则局限于文档级的关系,忽略了词一级的跨语言联系。这种方法在文档级(句子级)的文本分类中取得了不错的效果,但在属性级情感分类这种可能存在同句不同情感倾向的任务中,文档级的关系则显得比较粗糙,因而我们从属性级情感分类的特点出发,聚焦于更加细分的词一级跨语言关系,提出了基于跨语言图神经网络的属性级情感分类模型。

2 基于跨语言图神经网络的属性级情感分类模型

下面介绍我们提出的跨语言图神经网络的属性级情感分类模型,主要包括跨语言词-词、词-句图的构建和模型的整体结构。而在流程方面,首先,为了构建边密集图,我们将数据中目标语言数据和源语言数据翻译成彼此,再将翻译数据与原生数据中包括属性词在内的所有的词抽象成图中的一个词节点;而上下文与属性词则通过机器阅读理解的形式抽象成图中的句节点,使用一个双语词典构建包括双语词-词翻译边、双语词-句共现边以及单语词-句共现边在内的词级别跨语言图;随后使用机器阅读理解的形式组织上下文和属性词,并使用多语言预训练语言模型对其编码,将编码结果和跨语言图投入到图卷积神经网络 GCN 中。最后,GCN 将会输出对数据集中所有目标语言句节点及其翻译句节点的预测结果,并使用二者的分类概率联合判断最终结果。图 2 展示了我们提出的跨语言图神经网络的属性级情感分类模型整体的操作流程。



图 2 模型总体操作流程

图 3 则展示了我们所提模型的整体结构。

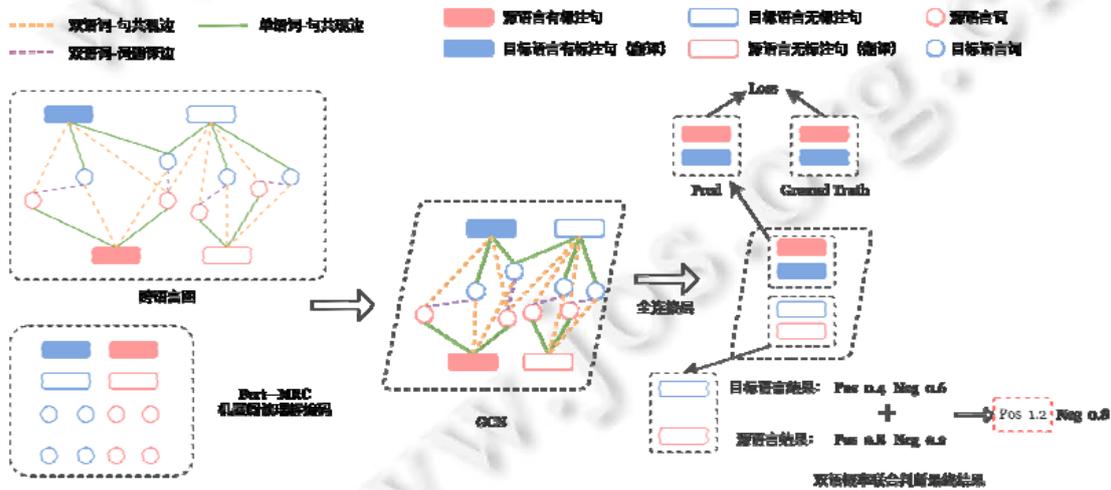


图 3 模型总体结构图

2.1 翻译数据

为了构建边密集跨语言图,我们对于所有数据均使用机器翻译将其由源语言数据翻译成目标语言数据(或与之相反)。为了避免机器翻译的质量影响到关键的属性词,我们将属性词和上下文分开翻译,然后再将翻

译过后的二者结合起来. 实际测试中发现, 采用加入机器翻译语料的方式可以有效地提升模型的性能表现. 具体方式如图 4 所示.

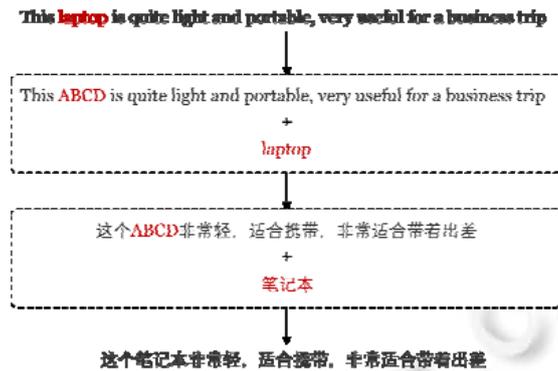


图 4 属性词分离翻译示意图

2.2 文档与单词的MRC编码

早期在属性级情感分类任务上用于区分上下文和属性词的方式主要包括 TD^[1]和 TC^[1], 这些方式在 RNN-base 的模型上取得了不错的性能表现. 而在预训练语言模型兴起后, 机器阅读理解(machine reading comprehension, MRC)形式逐渐取代了前两者. 因此, 我们在对属性词和上下文编码时, 采用机器阅读理解的形式来构成跨语言图中的句节点: 将需要判断情感倾向的属性词作为 Query, 将上下文作为 Context, 两者之间使用“[SEP]”符号隔开并投入到 XLM-R 模型中; 随后, 使用处于输入序列第 1 位的“[CLS]”的编码作为整个属性词和上下文的编码向量, 与建立的跨语言图一同投入到 GCN 中. 需要注意: 当一条数据包含 N 个属性词时, 将会由该条数据衍生出 N 个句节点, 这些句节点之间仅以 Query 部分进行区分, 具体形式如图 5 所示.

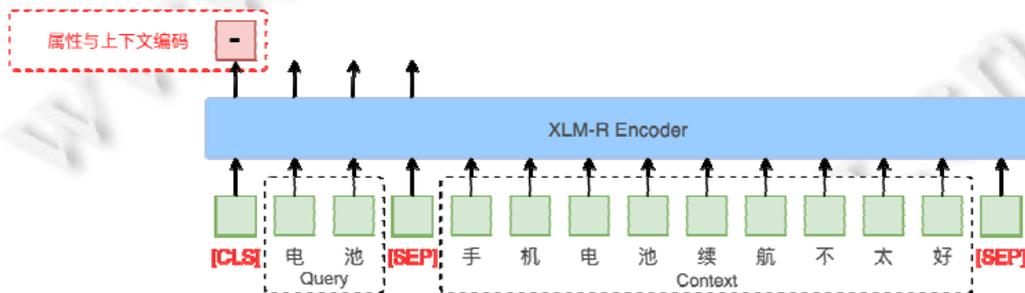


图 5 机器阅读理解形式编码示意图

2.3 跨语言图的构建

我们借鉴了其他一些在文本分类任务上使用图神经网络的方法^[4], 结合了跨语言任务的实际情况, 采取如下的流程建立投入 GCN 的跨语言图: 首先构建 6 种词句节点, 包括源语言句节点、源语言翻译句节点、目标语言句节点、目标语言翻译节点、源语言词节点、目标语言词节点, 其中, 所有的源语言节点属于有标注的训练集, 而目标语言节点则属于测试集, 在图中的以上 6 种节点之间, 我们建立以下 3 种边来在源语言和目标语言之间建立跨语言的语义联系(如图 6 所示).

- 单语言词-句共现边: 我们采取 TextGCN 中提出的建边方式, 构建了单语言词-句共现边来在单语言范围内词节点和句节点之间建立语义联系: 如果存在某一个词节点所代表的词存在于某一个句节点所代表的句中, 就在这两个节点之间建边. 不同的是: 在跨语言环境中, 该词节点单词可能不仅存在于其原生语言句节点中, 也可能存在于翻译语言句节点中, 例如: “书”不仅存在于原生中文句子“借

了一本书”中, 也存在于“*I like this book*”的翻译“我喜欢这部书”中。

- 双语言词-句共现边: 为了构建目标语和源语言之间的跨语言语义联系, 我们在单语言词-句共现边的基础上构建双语言的词句共现边, 主要包括两种建边方式。
 - 按词对齐(by words): 如若某个词节点代表的词的翻译存在于某个句节点所代表的句子中, 则在这个词节点和句节点之间建边。需要注意的是, 这个词节点和句节点所代表的文本应当属于不同的语言。例如: “书”的翻译“book”存在于句子“*I like this book*”中, 则在“书”和“*I like this book*”之间建边。
 - 按句对齐(by sentences): 如若存在一条原生句子 X_T 和另外一条与其对应的翻译句子 X_S , 则 X_T 中所有词的词节点与代表 X_S 的句节点相连接; 同理, X_S 中所有词的词节点与代表 X_T 的句节点相连接。
- 双语言词-词翻译边: 我们摒弃了句子级常见的句-句翻译边和基于余弦相似度的句-句相似边, 因为属性级情感分类任务的分类目标为方面级而非句子级, 因而经常存在同一个句子中的两个属性情感倾向相反的情况。例如图 1 中的“手机性能蛮好, 但是电池不耐用”, 若使用句-句相似边, 那么在对“性能”和“电池”的分类时会将两者相连(两者所属句子相同, 余弦相似度为 1), 然而两者的情感倾向却截然不同。句-句翻译边则也类似, 会将针对“性能”这一积极倾向属性的句子“手机性能蛮好, 但是电池不耐用”与针对“battery”这一消极倾向属性的句子“The performance is great, but the battery life is short”连起来。所以, 针对属性级情感分类任务的特点, 我们将翻译关系细分到词一级, 即若两个词节点所代表的词存在翻译关系时, 在二者之间建边。例如: “书”的翻译为“book”, 则在分别代表这两个词的词节点之间建边。



图 6 构建跨语言图示意图

2.4 图神经网络层

(1) 图卷积网络

卷积神经网络(convolution neural networks, CNN)在自然语言处理、图像处理等多个领域展现了优秀的建模效果。传统卷积神经网络模型的实质是在数据结构平移不变性的前提下进行特征提取, 因此相关应用只能局限在规则的欧氏空间中, 但现实中的许多数据是以不规则的非欧式空间结构呈现的, 比如分子结构数据、社交网络结构以及对话数据结构等。

Bruna 等人^[30]首次提出了将卷积神经网络模型泛化至非欧式空间的方法。之后的相关研究基本沿袭了卷积神经网络的网络局部连接和卷积核参数共享的核心思想, 通过在不规则的图结构上定义卷积算子、池化算子来提取空间特征。许多文献广泛使用了切比雪夫多项式来拟合卷积核, 在此基础上, Kipf 和 Welling^[31]一同

提出了图卷积网络模型(graph convolution network, GCN), 借助图的拉普拉斯矩阵的特征值和特征向量来定义卷积算子的方法. GCN 层之间的传播公式如下:

$$f(H^l, A) = \sigma \left(\hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} H^l W^l \right),$$

其中, H^l 为第 l 层输入特征, 对于输入层的话, H 就是输入节点编码; A 为图的邻接矩阵, \hat{A} 为 A 加上单位矩阵; \hat{D} 为 \hat{A} 度矩阵; W^l 为第 l 层参数矩阵; σ 为非线性激活函数.

(2) 图注意力网络

除去 GCN 外, 我们还使用了 Bengio 等人^[32]提出的图注意力机制(graph attention networks, GAT). 与上述 GCN 相比, GAT 与 GCN 都是将邻居顶点的特征聚合到中心顶点上, 利用图中的 local stationary 学习新的顶点特征表达. 不同的是: GCN 利用了拉普拉斯矩阵; 而 GAT 则利用注意力机制来强调输入的图中任意一节点 X_i 与其多个邻接节点之间关系的差异性, 从而将模型的注意力更多地分配给对学习当前节点特征更有帮助的邻接节点. GAT 通过归一化节点之间的相关系数来计算注意力系数, 随后, 根据注意力系数对特征加权求和得到新的节点特征, 其相关系数公式如下:

$$e_{ij} = a(\text{concat}(Wh_i, Wh_j)), j \in N_i,$$

其中, N_i 为节点 i 的邻接节点集合, e_{ij} 为节点 i 与其邻接节点 $j \in N_i$ 的相关系数 W 与 a 为可学习的参数. 计算出相关系数后, 通过归一化来计算注意力系数, 公式如下:

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(e_{ij}))}{\sum_{k \in N_i} \exp(\text{LeakyReLU}(e_{ik}))},$$

其中, α_{ij} 为节点 i 与其邻接节点 j 之间的注意力系数, LeakyReLU 为激活函数. 随后, 通过加权求和得到新的节点特征, 公式如下:

$$h_i' = \sigma \left(\sum_{j \in N_i} \alpha_{ij} Wh_j \right),$$

其中, σ 为非线性激活函数. h_i' 就是 GAT 输出的对于每个顶点 h_i 融合了邻域信息的新特征.

对于一组由源语言句 S_i^s 、目标语言句 S_i^t 、源语言词 W_i^s 、目标语言词 W_i^t 组成的跨语言属性级分类数据集 *vertex*, 参照之前设计的 3 种跨语言关联结构, 本文设定了如下 3 种具体的关系图结构.

• 单语言词-句共现边

在单语言范围内, 若存在源语言词 W_i^s 存在于源语言句 S_i^s 中, 或目标语言词 W_i^t 存在于源语言词 W_i^s 中, 则将二者之间相连, 权重默认设置为 1, 即无权图:

$$\text{edge}(S_i^s, W_j^s) = \begin{cases} 1, & \text{if } W_j^s \text{ in } S_i^s \\ 0, & \text{else} \end{cases}$$

或

$$\text{edge}(S_i^t, W_j^t) = \begin{cases} 1, & \text{if } W_j^t \text{ in } S_i^t \\ 0, & \text{else} \end{cases}$$

• 双语言词-句共现边

在跨语言范围内, 若存在源语言词 W_j^s 的翻译 $W_k^t = \text{translate}(W_j^s)$ 存在于目标语言句 S_i^t 中, 或目标语言词 W_j^t 的翻译 $W_k^s = \text{translate}(W_j^t)$ 存在于源语言词 S_i^s 中, 则将二者之间相连, 权重默认设置为 1, 即无权图:

$$\text{edge}(S_i^s, W_j^t) = \begin{cases} 1, & \text{if } \text{tranlate}(W_j^t) \text{ in } S_i^s \\ 0, & \text{else} \end{cases}$$

或

$$edge(S_i^s, W_j^s) = \begin{cases} 1, & \text{if } translate(W_j^s) \text{ in } S_i^s \\ 0, & \text{else} \end{cases}$$

- 双语言词-词翻译边

在跨语言范围内, 若存在源语言词 W_j^s 的翻译 $W_k^t = translate(W_j^s)$, 或目标语言词 W_j^t 的翻译 $W_k^s = translate(W_j^t)$, 则将二者之间相连, 权重默认设置为 1, 即无权图:

$$edge(W_j^s, W_k^t) = \begin{cases} 1, & \text{if } translate(W_k^t) \text{ in } W_j^s \\ 0, & \text{else} \end{cases}$$

通过以上 3 种边, 构建了我们跨语言属性级情感分类模型中的图结构 $G=(vertex, edge)$. 在我们的模型中, 图结构 G 与节点机器阅读理解编码结果 X_{vertex} 一同投入到图模型中, 输出结果再经过全连接层后进行分类.

2.5 模型训练与测试

给定上下文句子以及其中一个属性词, 我们的任务是预测属性词的二分类正负情感倾向. 我们使用了 Adam 优化器来训练模型, m 是类别数量; n 为样本数量; i 代表样本 i ; y_{ic} 为符号函数(0 或 1), 如果样本 i 的真实类别等于 c , 取 1, 否则取 0; p_{ic} 观测样本 i 属于类别 c 的预测概率, 我们的训练目标是 minimized 交叉熵损失函数来学习模型参数 θ , 损失函数定义如下:

$$L(\theta) = -\frac{1}{n} \sum_{i=0}^n \sum_{c=1}^m y_{ic} \log(p_{ic}; \theta).$$

由于我们的方法需要使用所有的样本建图, 包括训练集和测试集, 因此在训练时, 是将所有的样本均投入到模型中, 使用模型输出中的训练集结果来计算损失函数, 使用测试集结果来评估模型性能.

2.6 双语概率联合判断

我们在将数据投入到模型中之前, 会将目标语言和源语言数据分别机器翻译成彼此, 有标注的数据和无标注的数据会一起建图后投入到模型中. 输出后, 有标注的数据会计算 Loss 后反向传播; 而无标注的数据中, 每一条数据的源语言和目标语言结果将会对应相加后成为该条数据的最终判断结果. 即: 对于每一条未标注的目标语言数据 X_T , 会有一条与其对应翻译成源语言的未标注数据 X_S . 在模型测试时, 会分别输出 $P(X_T) = (p_{pos}^T, p_{neg}^T)$ 和 $P(X_S) = (p_{pos}^S, p_{neg}^S)$, 则在最终计算数据 X 的情感倾向 $T(X)$ 时, 采用:

$$T(X) = \max(p_{pos}^T + p_{pos}^S, p_{neg}^T + p_{neg}^S).$$

在实际测试中我们发现: 采用双语联合概率判断, 可以有效地提升模型的性能表现.

3 实验设置与结果分析

3.1 数据集

我们使用英文和中文分别作为源语言和目标语言进行实验. 在英文上, 我们使用 Semeval-14-Restaurant/Laptop Review 数据集. Semeval-14 是非常成熟的英文有标注属性级情感分析数据集, 主要是针对餐厅和笔记本电脑的电商评论内容, 其中每条数据均有一个或者多个由 {属性词, 情感词, 情感倾向} 组成的三元组. 我们对 Semeval-14 数据集进行筛选, 选择了三元组数量小于等于 3 的数据, 在 Res 和 Lap 两个领域上各随机选择了 300 条数据, 即一共 600 条数据, 且正负平衡. 在中文上, 我们使用淘宝的中文数据集, 有家具、百货、厨房、家纺、玩具共 5 个领域的的数据. 与英文数据类似, 我们在每个领域上同样筛选了正负平衡的 600 条数据.

在正向实验中, 我们使用英文数据作为训练集, 测试其在中文数据集的 5 个领域(每个领域 600 条)上各自的表现. 此外, 出于验证的目的, 我们也进行逆向的实验, 将中文作为源语言、英文作为目标语言进行测试. 将中文数据的 5 个领域混合后随机选择 600 条数据作为训练集, 分别测试在英文 Res 和 Lap 两个领域各 600 条的表现. 具体数据见表 1.

表 1 实验所用数据统计

数据集语言	分割	领域	数量
英文-中文	Train	Semeval-14 Res+Semeval-14 Lap	600
	Test	Jiaju	600
		Baihuo	600
		Chufang	600
		Wanju	600
Jiafang	600		
中文-英文	Train	Jiaju+Baihuo+Chufang+Wanju+Jiafang	600
	Test	Semeval-14 Res Semeval-14 Lap	600 600

3.2 评测指标与参数设置

我们使用 Huggingface Transformers 中的 XLM-RoBerta 作为机器阅读理解的编码器, 使用源语言数据(在我们的实验中为英文)微调后的 xlm-roberta-base 预训练参数对源语言和目标语言进行编码. 此外, 出于对比的目的, 我们也使用了没有进行机器翻译的训练集和测试集投入到模型中进行测试.

我们使用 $F1$ 值来评估模型性能, 取值范围为 0.0–1.0, 取值越高, 说明模型的性能越高:

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

经过调试测定, 我们最终设定 GCN 的隐藏层维度为 768, 输出维度为 256, 优化器为 Adam, 学习率为 0.000 1, $L2$ 正则为 0.001, 句子的最大长度为 130. xlm-roberta-base 预训练参数在英文数据上预先以 batchsize 为 32, 学习率为 0.000 1, 微调了 3 轮. 所有实验在 1 张 Nvidia RTX 2080S GPU 上完成.

3.3 基准模型介绍

为了评估模型效果, 我们选取了 4 种基准模型作为对比, 分别是 mBert、XLM、XLM-RoBerta、CLHG, 以下是对这些基准模型的简要介绍.

- multilingual BERT (mBert): 用 104 种语言组成的单一语料库训练的单个多语言 BERT 模型, 它的训练语料包含了 104 种语言的维基百科页面, 并且共享了一个词汇表. 多语言 BERT 在训练时既没有使用任何输入数据的语言标注, 也没有使用任何翻译机制来计算对应语言的表示. 多语言 BERT 对中文使用基于字符的标记化, 对所有其他语言, 使用 WordPiece 标记化;
- XLM: 尽管 mBert 的训练语料中包含超过 100 种语言, 但其模型本身并没有针对多语言进行优化: 大多数词汇没有在语言间共享, 因此能学到的跨语言知识是很有限的. 针对上述问题, XLM 的模型输入改用了字节对编码(BPE), 把输入按所有语言中最常见的词片段(sub-words)进行切分, 以此来增加跨语言共享的词汇. 此外, XLM 还提出了翻译语言模型(TLM): 在训练语料中加入含义相同语言不同的训练样本和位置编码, 训练中, 用一个语言的上下文信息去预测另一个语言被遮住的 token. 训练时, 则同时训练掩码语言模型(MLM)和 TLM;
- XLM-RoBerta: XLM-R 在方法上使用 XLM+RoBerta, 并增加了语种数量和训练数据集的数量, 使用超过 2 TB 的已经预处理过的 CommonCrawl 数据集, 以自监督的方式训练跨语言表征, 最终超过了单语言 BERT 模型在每种语言上的性能;
- CLHG: 使用了 XLM-R 作编码, 并在目标语言和源语言之间构建了跨语言图, 包括双语句-句翻译边、单语句-句相似边、单语词-句词性共现边等, 随后将编码信息和跨语言图通过图模型来输出分类结果.

3.4 实验结果与分析

(1) 与基准系统比较

表 2 列出了我们的模型与上述基准模型的实验结果. 其中, mBert、XLM、XLM-R 属于不使用翻译的方法, 它们均使用原生英文数据作训练集, 原生中文数据作测试集, 其 CN 列是指在原生中文数据测试集上的性能表现; 而 XLM-R (Translate)、CLHG 和我们提出的模型属于使用翻译的方法, 训练集为原生英文数据和由其

翻译来的中文数据, 测试集为原生中文数据和由其翻译来的英文数据, 其 CN 列是指在测试集中原生中文数据上的性能, EN 列是指在测试集中翻译英文数据上的性能, UNI 列则表示两者的概率联合判断.

表 2 英文数据集预测中文数据集结果

Method	Jiaju			Baihuo			Chufang		
	CN	EN	UNI	CN	EN	UNI	CN	EN	UNI
mBert	73.07			66.77			71.23		
XLM	73.82			73.2			71.08		
XLM-R	76.62			72.8			72.15		
XLM-R (translate)	74.87	70.08	75.86	72.16	66.82	70.66	70.39	67.82	70.7
CLHG	85.99	84.66	86.83	76.66	75.45	78.49	76.77	79.16	79.33
Ours-GAT (by sentence)	83.32	83.32	83.32	81.23	81.23	81.23	76.89	76.89	76.89
Ours-GAT (by words)	85.92	87.23	87.33	81.03	78.26	82.18	77.9	79.13	80.32
Ours-GCN (by sentence)	84.99	84.99	84.99	80.62	80.62	80.62	78.48	78.48	78.48
Ours-GCN (by words)	87.83	86.99	88.49	80.33	77.16	81.49	78.61	80.82	82.64

表 2 英文数据集预测中文数据集结果(续)

Method	Jiafang			Wanju			Avg.		
	CN	EN	UNI	CN	EN	UNI	CN	EN	UNI
mBert	65.82			64.24			68.23		
XLM	71.06			70.53			71.93		
XLM-R	74.28			76.49			74.46		
XLM-R (translate)	74.32	69.06	73.95	75.98	71.16	75.14	73.54	68.98	73.26
CLHG	77.48	73.49	78.79	81.49	81.64	82.32	79.67	78.88	81.15
Ours-GAT (by sentence)	79.82	79.82	79.82	82.39	82.38	82.42	80.73	80.72	80.73
Ours-GAT (by words)	77.29	75.28	77.12	81.27	81.9	83.32	80.68	80.36	82.05
Ours-GCN (by sentence)	80.08	80.08	80.08	83.16	83.32	83.32	81.46	81.49	81.49
Ours-GCN (by words)	79.33	76.77	78.29	82.83	82.49	84.16	81.78	80.84	83.01

从表中可以看出: 在使用 F1 指标的测评中, 我们的最终模型在使用英文数据集预测中文数据集的结果上取得了优于其他所有模型的效果(83.01). 其中, 双语言词-句共现边按词对齐(by words)的方式在两个语言单独的性能指标上与按句对齐(by sentences)互有高低, 但却可以通过联合概率判断的方式超越了按句对齐的性能指标, 而后者因为在互为翻译的两个句节点之间拥有数量更多的边, 导致模型在两个语言上的判断结果十分接近, 无法通过联合概率判断来提升性能表现. CLHG 中使用的句-句翻译边和句-句相似边过于强调句子级别的关系, 并不适合于属性级情感分类这种聚焦于词级别的文本分类任务, 在我们的实验中, 各项指标均低于我们提出的聚焦于词-词关系的模型.

此外, 由表中可以看出, 使用 GAT 的方法要显著地低于使用 GCN 的方法. 我们认为: GAT 虽然相较于 GCN 摒弃了拉普拉斯矩阵, 转而使用利用注意力机制动态地学习节点之间的关系, 但是这种方法并不适用于我们所提出的词-句建图的方式. 因为在我们所提出的图中的边是依据数据中的共现以及翻译这两种强关联关系构建的, 摒弃了由此而构建的拉普拉斯矩阵就忽略了节点之间本身存在的关系, 从而导致性能下降.

在基准模型中, 使用机器翻译的模型 XLM-R (Translate)显著地低于不使用机器翻译的模型 XLM-R, 我们认为, 这是由于机器翻译的质量所致.

我们还进行了逆向的实验, 即使用中文作为源语言, 英文作为目标语言, 结果见表 3.

表 3 逆向使用中文数据集预测英文数据集结果

Method	14 Laptop			14 Restaurant			Avg.		
	CN	EN	UNI	CN	EN	UNI	CN	EN	UNI
mBert	77.69	73.64	75.2	75.48	69.49	72.61	76.59	71.57	73.91
XLM-R (translate)	81.13	72.6	77.47	81.4	68.44	73.78	81.26	70.52	75.62
CLHG	84.15	80.43	85.04	78.33	76.89	78.92	81.24	78.66	81.98
Ours-GCN (by sentence)	86.07	85.91	85.99	83.91	83.83	83.91	84.99	84.87	84.95
Ours-GCN (by word)	85.23	81.67	85.52	77.57	78.64	80.32	81.4	80.15	82.92

由表 3 可见: 在逆向实验中, 我们的最终模型依然取得了优于基线模型的效果, 这说明我们的基于图神经网络的跨语言模型是有效的. 与正向实验中按句对齐和按词对齐在性能上较为接近不同, 逆向实验中按句

对齐显著地高于按词对齐, 分析发现: 在我们构建的图中, 与中文淘宝数据相连的边远多于与英文 SemEval 数据相连的边, 大概为 2:1, 当英文 SemEval 数据作测试集时, 由于按词对齐方式构建的边的数量更加稀疏, 导致与按句对齐方式的性能差距被放大.

(2) 不同影响因素的比较

在实验中, 我们根据该数据集的信息构成, 考虑了前述的多种图结构设计方法, 并对多种方法及其组合进行了效果评测, 用以探究边的类型对模型效果的影响. 从表 4 的实验结果中我们可以得到如下结论: 第一, 构建双语言词-词翻译边带来的提升最大(由 80.84 提升到 82.28), 这说明我们摒弃了 CLHG 使用的文档级(句子级)翻译关系, 将双语间语义关系细分到词一级的思路是有效的; 第二, 将图结构进行叠加可以取得进一步的提升, 如果忽视联合概率判断的性能表现的话, 显然在我们构建的跨语言图中, 边的数量越密集, 所带来的性能的提升也越大. 这说明我们对话结构的多种建模方法是有意义的, 增加边结构的数目能够更好地构建双语数据之间的语义联系.

表 4 不同图结构对模型效果的影响

边类别	XLM-R			1			2		
单语言词-句共现边	-			✓			✓		
双语言词-词翻译边	-			-			✓		
双语言词-句共现(按句对齐)	-			-			-		
双语言词-句共现(按词对齐)	-			-			-		
领域	CN	EN	UNI	CN	EN	UNI	CN	EN	UNI
Jiaju	74.87	70.08	75.86	84.64	84.66	86.16	84.97	85.15	88.16
Baihuo	72.16	66.82	70.66	78.33	76.16	79.83	78.64	77.76	80.49
Chufang	70.39	67.82	70.7	77.22	79.08	80.3	78.48	77.31	81.82
Jiafang	74.32	69.06	73.95	79.18	74.85	77.11	79.66	75.84	79.11
Wanju	75.98	71.16	75.14	81.48	78.99	80.81	81.3	79.32	81.83
AVG	73.54	68.98	73.26	80.17	78.74	80.84	80.61	79.07	82.28

表 4 不同图结构对模型效果的影响(续)

边类别	3			Full model		
单语言词-句共现边	✓			✓		
双语言词-词翻译边	✓			✓		
双语言词-句共现(按句对齐)	✓			-		
双语言词-句共现(按词对齐)	-			✓		
领域	CN	EN	UNI	CN	EN	UNI
Jiaju	84.99	84.99	84.99	87.83	86.99	88.49
Baihuo	80.62	80.62	80.62	80.33	77.16	81.49
Chufang	78.48	78.48	78.48	78.61	80.82	82.64
Jiafang	80.08	80.08	80.08	79.33	76.77	78.29
Wanju	83.16	83.32	83.32	82.83	82.49	84.16
AVG	81.46	81.49	81.49	81.78	80.84	83.01

除去探究边的类型对模型效果的影响之外, 我们也探究了加入机器翻译语料后带来的变化对于模型性能的影响, 在我们的模型中去除了翻译的句节点并与原本的模型性能作对比. 见表 5.

表 5 翻译语料对模型效果的影响

Method	Jiaju			Baihuo			Chufang		
	CN	EN	UNI	CN	EN	UNI	CN	EN	UNI
XLM-R	76.62	-	-	72.8	-	-	72.15	-	-
XLM-R (translate)	74.87	70.08	75.86	72.16	66.82	70.66	70.39	67.82	70.7
Ours-GCN (no translate)	83.59	-	-	77.74	-	-	76.79	-	-
Ours-GCN	87.83	86.99	88.49	80.33	77.16	81.49	78.61	80.82	82.64

表 5 翻译语料对模型效果的影响(续)

Method	Jiafang			Wanju			Avg.		
	CN	EN	UNI	CN	EN	UNI	CN	EN	UNI
XLM-R	74.28	-	-	76.49	-	-	74.46	-	-
XLM-R (translate)	74.32	69.06	73.95	75.98	71.16	75.14	73.54	68.98	73.26
Ours-GCN (no translate)	77.81	-	-	80.28	-	-	79.24	-	-
Ours-GCN	79.33	76.77	78.29	82.83	82.49	84.16	81.78	80.84	83.01

由表 5 可知: 我们通过加入机器翻译句节点的方法, 有效地提高了所提模型的性能表现(79.24 提升到 81.78)。我们认为, 这是因为加入机器翻译句节点有效地提升了所构建的跨语言图中边的数量。如表 6 所示: 加入翻译句节点将平均边数量由 133 031 提升到 219 816, 从而形成了边更为密集的跨语言图, 并且加入机器翻译句节点还可以通过双语言联合判断的方法继续提高模型性能。这说明, 我们构建边密集跨语言图的思路是正确的, 更多的边有效地构建了目标语言和跨语言数据之间的语义联系。

表 6 模型跨语言图中边的数量

边数量	Jiaju	Baihuo	Chufang	Jiafang	Wanju	Avg.
Ours	270 339	215 363	213 966	198 723	200 691	219 816
Ours (no translate)	162 022	129 778	130 978	121 112	121 265	133 031

(3) 实例分析

本文选择基线模型中表现较好的 CLHG 作为对照模型, 对图 5 中的实例进行分析。如图 7 例中所示: 由于属性级情感分类的特点, 经常存在某一句话中的两个属性词分布拥有不同的情感倾向, CLHG 中使用的文档级(句子级)翻译关系和相似关系会导致在构建跨语言图时, 在如同实例中这样两个同句不同倾向的数据之间建边, 从而导致模型的误判, 致使性能下降。本模型则从属性级情感分类的特点出发, 聚焦于词级别的语义联系, 成功理解了实例中的两个属性词进行分类。这说明, 本文使用的图结构能够更好地构建跨语言任务的语义联系, 从而在一定程度上提升了跨语言属性级情感分类的性能表现。

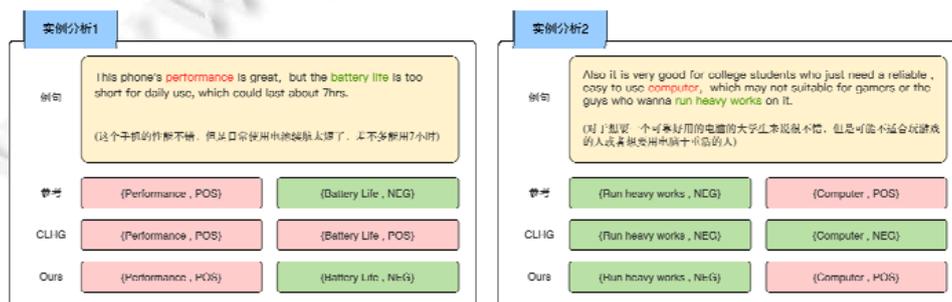


图 7 实例分析对照图

4 总 结

针对不同语言之间的数据资源分布不均匀的现象, 本文提出了一种基于跨语言图模型的属性级情感分类模型。该模型以预训练语言模型为框架, 通过图神经网络有效刻画了源语言和目标语言数据之间的语义关系, 增强了模型对词级别的跨语言语义的理解能力。我们在中英文跨语言数据集上, 分别测试了使用英语数据预测中文数据和逆向的使用中文数据预测英文数据的实验。结果表明, 我们的方法在 F1 指标上超出了其他基线模型。我们认为: 在跨语言属性级情感分类任务或者与其类似的其他一些聚焦于词级别的跨语言分类任务上, 都可以尝试通过使用构建词-词和词-句关系的思路来提升模型的性能表现。

我们下一步的研究方向主要是将跨语言建图的思路应用到属性级情感分析中更为复杂和困难的子任务中: 即属性词和情感词抽取。与属性级情感分类相比, 属性词和情感词抽取更加关注于每个句子中词级别的信息, 本文中提出的聚焦于词级别的跨语言关系的思路或许会在属性词抽取任务上带来更大的性能提升。

References:

- [1] Tang DY, Qin B, Feng XC, Liu T. Effective LSTMs for target-dependent sentiment classification. In: Proc. of the 26th Int'l Conf. on Computational Linguistics: Technical Papers (COLING 2016). 2016.
- [2] Wang YQ, Huang ML, Zhu XY, Zhao L. Attention-based LSTM for aspect-level sentiment classification. In: Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing. 2016.
- [3] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol.1 (Long and Short Papers). Minneapolis: Association for Computational Linguistics, 2019. 4171–4186.
- [4] Wang ZY, Liu X, Yang PJ, Liu SX, Wang ZS. Cross-lingual text classification with heterogeneous graph neural network. In: Proc. of the Association for Computational Linguistics 2021. 2021.
- [5] Yao L, Mao CS, Luo Y. Graph convolutional networks for text classification. In: Proc. of the AAAI Conf. on Artificial Intelligence, Vol.33. 2019. 7370–7377.
- [6] Song Y, Wang J, Jiang T, Liu Z, Rao Y. Targeted sentiment classification with attentional encoder network. In: Tetko I, Kůrková V, Karpov P, Theis F, eds. Proc. of the Artificial Neural Networks and Machine Learning: Text and Time Series (ICANN 2019). 2019.
- [7] Zeng BQ, Yang H, Xu RY, Zhou W, Han XL. LCF: A local context focus mechanism for aspect-based sentiment classification. Applied Sciences, 2019, 9(16): 3389.
- [8] Mao Y, Shen Y, Yu C, *et al.* A joint training dual-MRC framework for aspect based sentiment analysis. arXiv abs/2101.00816, 2021.
- [9] Wan XJ. Co-training for cross-lingual sentiment classification. In: Proc. of the Joint Conf. of the 47th Annual Meeting of the ACL and the 4th Int'l Joint Conf. on Natural Language Processing of the AFNLP. 2009. 235–243.
- [10] Andrade D, Sadamasa K, Tamura A, Tsuchida M. Cross-lingual text classification using topic-dependent word probabilities. In: Proc. of the 2015 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2015. 1466–1471.
- [11] Shi L, Mihalcea R, Tian MJ. Cross language text classification by model translation and semi-supervised learning. In: Proc. of the 2010 Conf. on Empirical Methods in Natural Language Processing. 2010. 1057–1067.
- [12] Li YG, Huang HY, Shi SM, Jian P, Su C. Title recognition of maximal-length noun phrase based on bilingual co-training. Ruan Jian Xue Bao/Journal of Software, 2015, 26(7): 1615–1625 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4630.htm> [doi: 10.13328/j.cnki.jos.004630]
- [13] Wang X, Shan LQ, Hou L, Yu JF, Chen J, Tao MY. Chinese English bilingual dictionary extraction based on encyclopedia corpus. Journal of Chinese Information Processing, 2021, 35(1): 25–33 (in Chinese with English abstract).
- [14] Zou WY, Socher R, Cer D, Manning CD. Bilingual word embeddings for phrase-based machine translation. In: Proc. of the 2013 Conf. on Empirical Methods in Natural Language Processing. 2013. 1393–1398.
- [15] Ziser Y, Reichart R. Deep pivot-based modeling for cross-language cross-domain transfer with minimal guidance. In: Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing. 2018. 238–249.
- [16] Peng XY, Zhou D. Survey of cross-lingual word embedding. Journal of Chinese Information Processing, 2020, 34(2): 1–15, 26 (in Chinese with English abstract).
- [17] Zhang MZ, Fujinuma Y, Graber JB. Exploiting cross-lingual subword similarities in low-resource document classification. In: Proc. of the AAAI Conf. on Artificial Intelligence, Vol.34. 2020. 9547–9554.
- [18] Clark K, Luong MT, Le QV, Manning CD. Electra: Pre-training text encoders as discriminators rather than generators. In: Proc. of the Int'l Conf. on Learning Representations. 2020.
- [19] Conneau A, Lample G. Cross-lingual language model pretraining. In: Advances in Neural Information Processing Systems. 2019. 7059–7069.
- [20] Conneau A, Khandelwal K, Goyal N, Chaudhary V, Wenzek G, Francisco, Guzmán, Grave E, Ott M, Zettlemoyer L, Stoyanov V. Unsupervised cross-lingual representation learning at scale. In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. 2020. 8440–8451.
- [21] Liu YH, Ott M, Goyal N, Du JF, Joshi M, Chen DQ, Levy O, Lewis M, Zettlemoyer L, Stoyanov V. RoBERTa: A robustly optimized BERT pretraining approach. 2019. <https://arxiv.org/pdf/1907.11692.pdf>

- [22] Fei HL, Li P. Cross-lingual un-supervised sentiment classification with multi-view transfer learning. In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. 2020. 5759–5771.
- [23] Chen Z, Qian TY. Transfer capsule network for aspect level sentiment classification. In: Proc. of the 57th Conf. of the Association for Computational Linguistics (ACL). Florence, 2019. 547–556.
- [24] Yao L, Mao CS, Luo Y. Graph convolutional networks for text classification. In: Proc. of the AAAI Conf. on Artificial Intelligence, Vol.33. 2019. 7370–7377.
- [25] Hu LM, Yang TC, Shi C, Ji HY, Li XL. Heterogeneous graph attention networks for semi-supervised short text classification. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: Association for Computational Linguistics, 2019. 4821–4830.
- [26] Ding KZ, Wang JL, Li JD, Li DC, Liu H. Be more with less: Hypergraph attention networks for inductive text classification. In: Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing (EMNLP). 2020. 4927–4936.
- [27] Zhang YF, Yu XL, Cui ZY, Wu S, Wen ZZ, Wang L. Every document owns its structure: Inductive text classification via graph neural networks. In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. 2020. 334–339.
- [28] Wang G, Li HY, Qiu YF, Yu BW, Liu TW. Aspect-based sentiment classification via memory graph convolutional network. Journal of Chinese Information Processing, 2021, 35(8): 98–106 (in Chinese with English abstract).
- [29] Li Z, Kumar M, Headden W, Yin B, Wei Y, Zhang Y, Yang Q. Learn to cross-lingual transfer with Meta graph learning across heterogeneous languages. In: Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing (EMNLP). 2020. 2290–2301.
- [30] Bruna J, Zaremba W, Szlam A, Lecun Y. Spectral networks and locally connected networks on graphs. In: Proc. of the 2nd Int'l Conf. of Learning Representations. 2014.
- [31] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: Proc. of the 5th Int'l Conf. on Learning Representations. 2017.
- [32] Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph attention networks. In: Proc. of the Int'l Conf. on Learning Representations. 2018.

附中文参考文献:

- [12] 李业刚, 黄河燕, 史树敏, 鉴萍, 苏超. 基于双语协同训练的最大名词短语识别研究. 软件学报, 2015, 26(7): 1615–1625. <http://www.jos.org.cn/1000-9825/4630.htm> [doi: 10.13328/j.cnki.jos.004630]
- [13] 王星, 单力秋, 侯磊, 于济凡, 陈吉, 陶明阳. 基于百科语料的中英文双语词典提取. 中文信息学报, 2021, 35(1): 25–33.
- [16] 彭晓娅, 周栋. 跨语言词向量研究综述. 中文信息学报, 2020, 34(2): 1–15, 26.
- [28] 王光, 李鸿宇, 邱云飞, 郁博文, 柳厅文. 基于图卷积记忆网络的方面级情感分类. 中文信息学报, 2021, 35(8): 98–106.



鲍小异(1997—), 男, 硕士生, 主要研究领域为自然语言处理, 情感分析.



王中卿(1987—), 男, 博士, 副教授, CCF 专业会员, 主要研究领域为自然语言处理, 情感分析.



姜晓彤(1997—), 女, 博士生, CCF 学生会员, 主要研究领域为自然语言处理.



周国栋(1967—), 男, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为自然语言处理.